

EXHIBIT E



Contents lists available at ScienceDirect

Mutation Research/Genetic Toxicology and Environmental Mutagenesis

journal homepage: www.elsevier.com/locate/gentox
 Community address: www.elsevier.com/locate/mutres



IWGT report on quantitative approaches to genotoxicity risk assessment I. Methods and metrics for defining exposure–response relationships and points of departure (PoDs)[☆]



James T. MacGregor^{a,*}, Roland Frötschl^b, Paul A. White^{c,**}, Kenny S. Crump^d,
 David A. Eastmond^e, Shoji Fukushima^f, Melanie Guérard^g, Makoto Hayashi^h,
 Lya G. Soeteman-Hernándezⁱ, Toshio Kasamatsu^j, Dan D. Levy^k, Takeshi Morita^l,
 Lutz Müller^m, Rita Schoenyⁿ, Maik J. Schuler^o, Véronique Thybaud^p, George E. Johnson^q

^a Toxicology Consulting Services, Bonita Springs, FL, USA

^b Bundesinstitut für Arzneimittel und Medizinprodukte, Bonn, Germany

^c Health Canada, Ottawa, ON, Canada

^d Ruston, LA, USA

^e University of California, Riverside, CA, USA

^f Japan Bioassay Research Center, Hadano, Kanagawa, Japan

^g F. Hoffmann-La Roche Ltd., Basel, Switzerland

^h Biosafety Research Center, Foods, Drugs and Pesticides, Iwata, Shizuoka, Japan

ⁱ RIVM-National Institute for Public Health and the Environment, Bilthoven, Netherlands

^j Kao Corporation, Ichikai-Machi, Haga-Gun, Tochigi, Japan

^k U.S. Food and Drug Administration, College Park, MD, USA

^l National Institute of Health Sciences, Tokyo, Japan

^m Hoffmann-La Roche Inc., Basel, Switzerland

ⁿ U.S. Environmental Protection Agency, Washington, DC, USA

^o Pfizer, Inc., Groton, CT, USA

^p Sanofi, Vitry sur Seine, France

^q Swansea University, Swansea, United Kingdom

ARTICLE INFO

Article history:

Received 25 September 2014

Accepted 29 September 2014

Available online 13 October 2014

Keywords:

Genotoxicity

Quantitative risk assessment

Points of departure

Benchmark dose

ABSTRACT

This report summarizes the discussion, conclusions, and points of consensus of the IWGT Working Group on Quantitative Approaches to Genetic Toxicology Risk Assessment (QWG) based on a meeting in Foz do Iguaçu, Brazil October 31–November 2, 2013. Topics addressed included (1) the need for quantitative dose–response analysis, (2) methods to analyze exposure–response relationships & derive point of departure (PoD) metrics, (3) points of departure (PoD) and mechanistic threshold considerations, (4) approaches to define exposure-related risks, (5) empirical relationships between genetic damage (mutation) and cancer, and (6) extrapolations across test systems and species. This report discusses the first three of these topics and a companion report discusses the latter three. The working group critically examined methods for determining point of departure metrics (PoDs) that could be used to estimate low-dose risk of genetic damage and from which extrapolation to acceptable exposure levels could be made using appropriate mode of action information and uncertainty factors. These included benchmark doses (BMDs) derived from fitting families of exponential models, the *No Observed Genotoxic Effect Level* (NOGEL), and “threshold” or breakpoint dose (BPD) levels derived from bilinear models when mechanistic data supported this approach. The QWG recognizes that scientific evidence suggests that thresholds below which genotoxic effects do not occur likely exist for both DNA-reactive and DNA-nonreactive substances, but notes that small increments of the spontaneous level cannot be unequivocally excluded either by

[☆] The opinions and recommendations expressed in this publication are those of the authors, and do not necessarily reflect those of the institutions with which they may be affiliated.

* Corresponding author. Tel.: +1 410 991 9948; fax: +1 239 947 7447.

** Corresponding author.

E-mail addresses: jtmacgregor@earthlink.net (J.T. MacGregor), paul.white@hc-sc.gc.ca (P.A. White).

experimental measurement or by mathematical modeling. Therefore, rather than debating the theoretical possibility of such low-dose effects, emphasis should be placed on determination of PoDs from which acceptable exposure levels can be determined by extrapolation using available mechanistic information and appropriate uncertainty factors. This approach places the focus on minimization of the genotoxic risk, which protects against the risk of the development of diseases resulting from the genetic damage. Based on analysis of the strengths and weaknesses of each method, the QWG concluded that the order of preference of PoD metrics is the statistical lower bound on the BMD > the NOGEL > a statistical lower bound on the BPD. A companion report discusses the use of these metrics in genotoxicity risk assessment, including scaling and uncertainty factors to be considered when extrapolating below the PoD and/or across test systems and to the human.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

The International Workshops on Genotoxicity Testing (IWGT) bring together experts in the field of genetic toxicology approximately every four years to develop consensus recommendations on genetic toxicology testing methods and testing strategies. Beginning in 2002 a series of working groups addressed different aspects of hazard identification/characterization and risk assessment strategies [1,2]. At the 2009 meeting in Basel, Switzerland, the working group on follow-up testing of agents that elicited positive responses in *in vivo* assays [3] agreed on “the need for a consensus about the most appropriate mathematical models and statistical analyses for defining non-linear dose–responses and exposure levels associated with acceptable risk”. Those points were recently further discussed by a Working Group on Quantitative Approaches to Genetic Toxicology Risk Assessment (QWG) organized as part of the October 31–November 2, 2013 meeting of the IWGT in Foz do Iguaçu, Brazil. This QWG addressed the following topics:

1. The need for quantitative dose–response analysis of genetic toxicology data
2. The existence and appropriate evaluation of threshold responses
3. Methods to analyze exposure–response relationships & derive points of departure (PoDs) for extrapolation to low-dose exposure levels
4. Approaches to define exposure-related risks
5. Empirical relationships between genetic damage (mutation) and cancer
6. Extrapolation across test systems and species.

This report is the first of two that summarize the discussions and recommendations of the QWG. It focuses on issues 1–3 listed above, provides the background and rationale for quantitative approaches in genotoxicity hazard identification, and addresses methods and metrics for defining dose–response relationships and determining PoDs for extrapolation to low-dose exposure levels. A second report will consider issues 4–6 and address considerations for extrapolation below PoDs, uncertainty factors in extrapolation from PoDs, extrapolation across test systems and species, and overall considerations on the use of quantitative approaches in genotoxicity risk assessment.

2. Background and objectives: the need for quantitative dose–response analysis of genetic toxicology data

Regulatory genetic toxicology testing was first introduced in North America, Europe and Japan in the 1970s and early 1980s [4–9]. These regulations and recommendations emphasized the use of a specified hazard screening battery consisting of *in vitro* tests for gene mutation and chromosomal damage and an *in vivo* test for chromosomal damage in hematopoietic tissue. These reports and

guidelines also recognized the need for quantitative *in vivo* data to establish quantitative estimates of human risk. In its first comprehensive guideline on the application of genetic testing procedures, the U.S. Department of Health Education, and Welfare (DHEW) Subcommittee on Environmental Mutagenesis (1974–1977) stated the following [10]:

- “It is not sufficient to merely identify substances which may pose a genetic hazard to the human population.”
- “...it is necessary to obtain quantitative data from relevant animal model systems from which extrapolations to humans can be made to predict virtually safe or tolerable levels of exposure”.

However, this early recommendation for a quantitative analysis of *in vivo* risk was eclipsed by the initial enthusiasm generated by the finding of an apparent close concordance between *in vitro* mutation assays and *in vivo* cancer outcomes, initiated in large part by the analysis of Ames and McCann [11,12]. Within a few years a perception developed that mutagens and carcinogens were relatively rare and that hazard screening could identify them so that human exposure could be limited or eliminated (*i.e.*, by application of the precautionary principle). At that time, cancer became the main health consequence of concern in many regulatory agencies, including the U.S. Food and Drug Administration (FDA); this was a change from the original DHEW report that emphasized the importance of heritable germ cell damage. Moreover, database analyses suggested that potency measured using *in vitro* tests was not a reliable indicator of *in vivo* potency [13,14] and the relative insensitivity of the *in vivo* tests [15,16] discouraged their use outside of a battery of tests designed for hazard identification. As a result, regulatory testing and decision-making have until recently been based largely on qualitative outcomes of *in vitro* and/or *in vivo* genotoxicity tests, leading to dichotomous classification of an agent as genotoxic/mutagenic/clastogenic/aneuploid, or not, with cancer as the principal genetic disease of concern.

More recently it has become recognized that qualitative outcomes of *in vitro* genotoxicity tests do not adequately correspond to the outcomes of rodent carcinogenicity bioassays or with *in vivo* genotoxicity. Among the key publications demonstrating the shortcomings of over-reliance on screening tests for genotoxic hazards is Snyder and Green's report [17] that ~25% of pharmaceuticals in the 1999 Physicians' Desk Reference (PDR) were positive in *in vitro* mammalian cell tests (after exclusion of chemotherapy agents) even though their use is considered safe. Moreover, Sofuni et al. [18] report that 38% of 1049 new chemical substances evaluated under the Chemical Substances Control Law in Japan were positive in a mammalian cell chromosomal aberration assay. Further, the extensive summary by Kirkland et al. [19], which concluded that *in vitro* cytogenetic tests exhibit high sensitivity but low specificity when compared with *in vivo* rodent cancer test outcomes (*i.e.*, with many positive outcomes for agents that were negative in

rodent cancer bioassays), demonstrates that *in vitro* assays often show effects that are not reproduced in *in vivo* studies. Some limitations of the Kirkland analysis have been pointed out [20], and more recent surveys suggest that current methods reduce the fraction of positive results [21,22]. Nonetheless, it is clear that cellular exposures attained in *in vitro* assays often greatly exceed those achievable *in vivo* and potentially lead to toxicity and cellular disturbances that cause genetic damage that would not occur under conditions more reflective of actual *in vivo* exposures. Because qualitative classification as positive or negative for a particular endpoint is often insufficient for risk assessment, regulatory agencies have already begun to mandate more comprehensive consideration of the relevance of *in vitro* positive findings, including differences in exposures that might influence *in vivo* outcomes. Therefore an overall weight of evidence (WoE) approach that takes into account genotoxicity data in the context of all other relevant information is currently being emphasized when evaluating genotoxic hazard and risk. Examples include the following:

- FDA guidance on integration of genetic toxicity data during review of preclinical drug safety studies [23], which emphasizes consideration of the overall weight of evidence
- the ICHS2 revision that places increased weight on *in vivo* outcomes [24]
- for genotoxic impurities, the EMA and draft ICH and FDA guidances that recognize both thresholds of toxicological concern (TTC) for genotoxic substances and exposure limits associated with negligible risk based on magnitude and duration of exposure [25–27]

Several consortia and working groups have also highlighted the need to include quantitative information (i.e., dose–response information) in the weight of evidence approach, and are currently engaged in the development of guidelines for quantitative approaches that will improve the assessment of potential risks from exposures to genotoxic agents. These include the IWGT (see, e.g., [3]), the Health and Environmental Sciences Institute of the International Life Sciences Institute [28–30]), and other groups (e.g., [31]).

There was a clear consensus of the QWG that increased utilization of quantitative information, especially *in vivo* information, would improve the assessment of risks posed by exposure to genotoxic agents. However, there are currently no internationally-accepted guidelines for the quantitative analyses of genetic toxicity dose–response functions or for the use of quantitative potency metrics in risk assessment and regulatory decision-making. Therefore a principal objective of the QWG was the development of recommendations for appropriate methodologies to describe genotoxicity exposure–response relationships and to derive points of departure (PoD) metrics. An additional objective was to develop recommendations for the utilization of such metrics for the assessment of human risk associated with exposure to genotoxic agents.

In early discussions, the QWG agreed that it is not practical to conduct comprehensive quantitative risk assessments for all existing chemicals, due to the overwhelming cost and resources that would be needed to obtain the necessary data and to conduct such analyses. Therefore, there is a need for a tiered approach that depends to a large extent on the level of concern established by hazard screening, the anticipated level of exposure, and evaluation of other available data. Such strategies have been described elsewhere and were not a focus of the QWG discussions. Rather, the QWG focused on quantitative approaches to be used when the degree of exposure and potential hazard warranted a quantitative analysis to support risk characterization and risk management.

The risk assessment and risk characterization process for genotoxic agents has been described previously by a working group sponsored jointly by the United Nations Environmental Program (UNEP) and the International Commission for Protection Against Environmental Mutagens and Carcinogens (ICPEMC) [32,33], which was based on the framework for risk assessment previously defined by a National Academy of Sciences working group [34]. This process is illustrated in the figure below, redrawn, with permission, from the 1994 ICPEMC/UNEP publication [35] (Fig. 1).

The QWG focused on the dose(exposure)–response stage of this process, including the use of quantitative data from laboratory investigations of genotoxicity and related endpoints to estimate acceptable human exposures; *i.e.*, definition of exposure response relationships, determination of and extrapolation from points of departure, integration of exposure information with the exposure–response relationships to estimate responses at exposure levels relevant to those in humans, and integration of the available information with uncertainty factors to extrapolate to human health risk. Due to the diversity of products and environmental exposure scenarios that may require evaluation, it is not possible or desirable to recommend one set of criteria for risk management that applies to all situations. Therefore the QWG efforts focused on the identification of key elements and factors that should be used in a general framework for risk assessment that is applicable in most situations. Risk management is expected to be implemented by regulatory agency professionals and risk managers responsible for specific product classes and environmental exposure situations, and will incorporate specific criteria and acceptable risk levels applicable to those particular situations and exposure scenarios.

Examples of specific regulatory criteria and testing requirements that have already been defined for some product classes and environmental exposure situations include:

- FDA food additive levels of concern based on a combination of chemical structure class and exposure, [36].
- Thresholds of regulation: food exposures below which testing requirements are reduced [37] or eliminated [38].
- The exposure-related allowance for genotoxic impurities during drug development specified in the EMA, FDA, and ICH guidance documents [25–27], based on carcinogenicity potency data (*i.e.*, 10^{-5} or 10^{-6} increased risk of cancer).
- The Canadian Department of National Health and Welfare (CDHNW, now known as Health Canada) testing requirements for environmental contaminants, drugs, and foods, including categorization by level of concern [6,39].

2.1. The existence and evaluation of threshold responses

An important topic discussed by the QWG was the existence of thresholds for genotoxic effects. There is general consensus that genotoxic agents acting by some non-DNA-reactive mechanisms are expected to exhibit a threshold exposure level (*i.e.*, dose or concentration) below which no biological effect is expected. Examples of these mechanisms include mitotic spindle disturbance, nucleotide pool imbalance, glutathione depletion, and inhibition of DNA synthesis [28,40–42]. Additional examples and references are given in the QWG companion manuscript on the use of quantitative data in genetic toxicity risk assessment [43]. DNA-reactive genotoxins, in contrast, have often been considered to have a finite risk at any dose (e.g., [44]). However, regardless of the mechanism it is impossible to determine experimentally at low dose ranges whether or not there is a small incremental risk within the normal distribution of the ever-present spontaneous background. Further, a small increase in a unique deleterious genetic alteration that is not present in the normal spontaneous

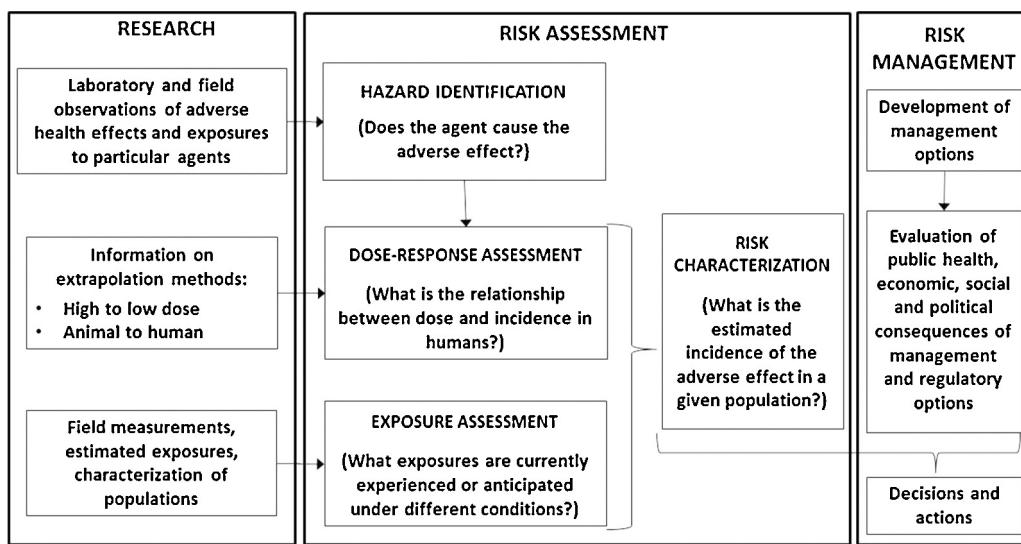


Fig. 1. Elements of risk assessment and risk management.

background would be very difficult to identify or rule out. Moreover, any set of dose-response data – no matter how extensive – will be mathematically consistent with both threshold and low-dose linear responses [45]. Since neither experimentation nor mathematical analysis can definitively exclude small genotoxic effects within the normal background range, the QWG concluded that it is not productive to argue about the existence of thresholds for mutagenic and clastogenic agents. Rather, emphasis should be placed on determination of PoDs from which acceptable exposure levels can be determined by extrapolation using available mechanistic information and appropriate uncertainty factors. This approach places the focus on minimization of the genotoxic risk, which is expected to protect against the risk of the development of diseases that could result from the genetic damage. Such an approach has been recommended by the Health Protection Branch of the Canadian Department of National Health and Welfare (now known as Health Canada) [6,39,46], which recognized “genotoxic activity as a potential hazard for adverse human health effects, and, accordingly, that genotoxicity is a *bona fide* toxicological endpoint”. Thus, in the late 1980s and early 1990s regulatory agencies such as Health Canada were already recognizing the need for a uniform methodology for the quantitative toxicological evaluation of substances that enhance the risk of adverse health effects *via* induction of genetic damage.

2.2. Methods to analyze exposure-response relationships & derive point of departure (PoD) metrics

The discussion that follows focuses on appropriate methods for defining exposure-response relationships for genetic toxicity and establishing PoDs from which extrapolation can be made to determine acceptable exposure limits. When defining approaches to derive appropriate PoDs, one should first identify desirable characteristics of each approach and metric. These desirable characteristics include:

- Well defined & robust:** the method and associated PoD metric should be applicable to a wide variety of dose-response data, including sparse data
- Conservative:** the method should employ approaches that account for data uncertainty and attempt to avoid

underestimating the actual risk by defining the PoD as a statistical lower bound

- Transparent:** the assumptions required to derive the PoD should be clearly articulated
- Ease of calculation:** the procedure for analysis and PoD determination should be straightforward, using publicly available methods and/or software
- Interpretable biological meaning:** the result should be linked to undesirable physiologic effects that are in turn related to human disease. The PoD on these results should be an index of risk in the lower measurable range.

Three main approaches have been used to date to derive PoDs for human health risk assessment (HHRA).

- The No Observed Genotoxic Effect Level (NOGEL), which is the no-observed effect level (NOEL) for a genotoxicity endpoint, is defined as the highest tested dose for which no statistically significant increase in the incidence of the genotoxic effect is observed relative to an appropriate untreated or vehicle-treated control, and below which there are no statistically significant increases in the genotoxic effect. The NOEL and NOAEL (no observed adverse effect level) are widely used in regulatory settings (e.g., [36,47]), and the NOGEL is directly analogous because genetic damage is generally considered to be an adverse effect.
- The benchmark dose (BMD), or benchmark concentration (BMC), is a dose or concentration that results in a defined change (benchmark response or BMR) in the level of an adverse response. Different benchmark approaches are generally used with quantal data (in which a subject is classified as affected or not affected, such as cancer occurrence) and continuous data (in which a subject's response can be essentially any value in a range, such as mutation or micronucleus frequency when the subject's response is considered to be a continuous variable). A statistical lower bound on the BMD (BMDL) can be used as the PoD [29,30,48–50]. The BMD approach has been used by many agencies for determining PoDs for health-based guidance (e.g., [51–53]).

When applied to quantal data the benchmark dose can be defined as the dose that results in an additional fraction (BMR) of subjects that have the adverse response. If $P(x)$ represents the probability of responding at a dose of x , then $P(0)$ is the

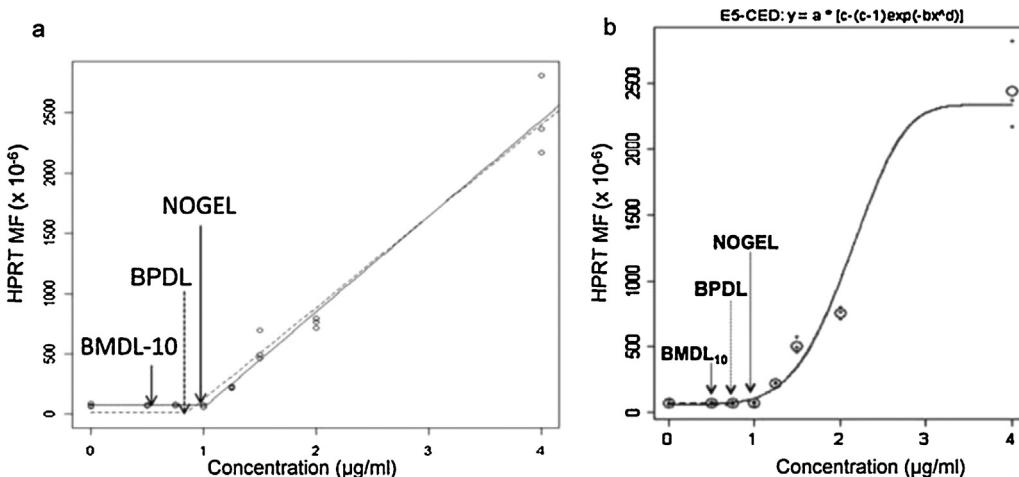


Fig. 2. Dose-response modeling results for HPRT gene mutations induced by MMS *in vitro* in AHH-1 cells. (a) Bilinear dose-response modeling used to define the BPD and BPDL, (b) BMD dose-response modeling used to define the BMD₁₀ and BMDL₁₀. The three parameters, including a NOGEL derived using a 1-sided Dunnett's test, a BPDL, and a BMDL₁₀, are shown in both graphs for comparison. For (a), the solid line shows the best fit of the bilinear function, and the dotted line shows the bilinear model with the BPDL as the lower 95% confidence limit of the breakpoint dose [64].

Reproduced, with permission, from Gollapudi et al. [29].

background response rate. Accordingly, $P(x) - P(0)$ represents the additional fraction responding at a dose of x , and the BMD is defined as the dose that satisfies

$$P(\text{BMD}) - P(0) = \text{BMR} \quad (1)$$

where BMR is a number between 0 and 1 selected in advance. If, for example, BMR is selected as $\text{BMR} = 0.1$, the corresponding BMD is often denoted by BMD_{10} as it is the dose that results in an additional 10% of subjects observed to be affected. To calculate a BMD one must posit a mathematical form for the probability of response, $P(x)$. Thus, $P(x)$ will involve statistical parameters whose values are estimated by fitting $P(x)$ to data. Statistical procedures can be used for evaluating how well the selected dose response fits the data and for calculating the BMDL (lower statistical bound on the BMD).

When applied to continuous data the benchmark dose can be defined as the dose that causes a specified average change (BMR) in a response in the direction that is considered adverse (either increasing or decreasing) relative to that response in unexposed subjects. If the posited mathematical form $F(x)$ represents the average level of response when exposed to dose x , then $F(x) - F(0)$ represents the average change in response caused by dose x . If an increase in the normal background response is considered adverse then the BMD is defined as the dose that satisfies

$$\frac{F(\text{BMD}) - F(0)}{F(0)} = \text{BMR} \quad (2)$$

If, instead a decrease in the normal background response is considered adverse, then the numerator is replaced by $F(0) - F(\text{BMD})$. Alternatively, the BMR for continuous data can be defined as a change in response relative to the normal variation as measured by σ , the standard deviation, in unexposed subjects,

$$\frac{F(\text{BMD}) - F(0)}{\sigma} = \text{BMR} \quad (3)$$

[54–56]. The BMD so-defined is sometimes denoted by BMD_{1SD} (the change in response corresponding to one standard deviation). A variation of this approach, called the “hybrid approach”, involves defining the most extreme responses (e.g., the most extreme 5%) among unexposed

subjects to be “affected”. This allows the BMD from continuous data to have the same interpretation as those from quantal data (*i.e.*, both defined by Eq. (1) (see [54,55,57–59] for more details)). Otherwise, BMDs defined from Eqs. (1), (2), or (3) may not be comparable even if based on the same numerical value of BMR. A majority of members of the QWG expressed a preference for defining BMD from continuous genotoxic data using Eq. (2) rather than Eq. (3).

The BMDL, a lower statistical confidence bound on the BMD, can be a suitable PoD for either continuous or quantal genotoxicity data. The BMDL satisfies the desirable characteristic of being conservative by incorporating a quantitative estimate of one of the types of model uncertainty.

When using the BMD approach, the functions, $P(x)$ or $F(x)$, used to model the incidence or continuous response as a function of dose, x , must be specified, and the criteria used to select the mathematical form of $P(x)$ or $F(x)$ need to be appropriately defined. For BMD analysis of continuous data, it is often reasonable to assume a log-normal distribution for the response at a fixed dose, where the variation (log-normal standard deviation) does not depend upon dose. If the variation does depend upon dose, it may be necessary to model that variation as a function of dose (*i.e.*, model σ as a function of dose, $\sigma = \sigma(x)$). A useful measure of uncertainty when using the BMD method is the BMDU/BMDL ratio (BMDU is the statistical upper limit on the BMD). A large BMDU/BMDL ratio suggests that the data may not be suitable for defining a sufficiently reliable PoD [60].

In the QWG analyses (Fig. 2), which involved continuous responses, a BMR of 10% was used, corresponding to an increase equal to 10% of the background (negative control) level. It should be noted that for quantal data, such as cancer incidence data, a BMR of 10% in absolute incidence rate is often selected to calculate a PoD. It should be emphasized that while both approaches generate a PoD that is referred to as a “ BMD_{10} ”, they are substantially different because the continuous response analysis, which is generally applied to genotoxicity data when the response of a subject is considered to be a continuous variable, is based on a BMR of a specified percentage increase (often 10%) of the spontaneous incidence, whereas the quantal analysis of cancer data is generally based on an absolute increase of 10% tumor incidence.

Table 1

Characteristics and recommended data needs of methods considered by the QWG.

	BMD	NOGEL	Bilinear
Recommended minimum data	3 doses plus control	3 doses plus control*	≥5 doses (including the negative control)
Amount data used	Whole data set used	Only data in one dose group is used	Whole data set used
Known effect size?	Yes (BMR)	No	No
Covariate analysis possible?	Yes	No	No
Estimate heavily dependent on experimental conditions?	No	Yes	Yes
Confidence limits	Yes	No	Yes

* Generally required to approximate NOGEL unless a limit dose is negative.

Software packages designed for performing the calculations needed for BMD analysis include PROAST [61] and the EPA's BMDS software [62]. Both are available for download and use without cost: (http://www.rivm.nl/en/Documents_and_publications/Scientific/Models/PROAST; <http://www.epa.gov/ncea/bmids/dwnldu.html>).

(iii) The threshold or breakpoint dose lower confidence bound (BPDL), determined using a bi-linear model, has also been used in human health risk assessment [63]. In this approach, a bi-linear dose response function is fit to the data: it is composed of two intersecting straight lines with the low-dose line having a zero slope. Examples of such models are those described by Lutz and Lutz [64] and Muggeo [65]. A statistical lower bound on the point of intersection of the two lines is considered an estimate of the threshold dose (highest dose with no change in response, or BPD) and can be used as a PoD. This method assumes the existence of a threshold dose, and consequently should not be used unless there is general agreement that the available mechanistic information supports this assumption. This approach is equivalent to a BMD approach that uses a specific dose response model (*i.e.*, a bi-linear model) and a BMR = 0. Bilinear modeling [29,64–66] can be used to reject a linear dose response relationship across the entire dose range. Additionally, when statistical assumptions are met and the analysis is supported by mechanistic information, bilinear modeling can be used to accept a dose–response relationship with zero slope below the estimated threshold or BreakPoint Dose (BPD).

The Quantitative Analysis Working Group of the Genetic Toxicology Technical Committee of the Health and Environmental Sciences Institute (part of the International Life Sciences Institute: ILSI-HESI GTTC QAW) guidance document and R packages contain statistical software for fitting the bilinear model [30,67]. Note that a log-transformation of the doses before fitting the bi-linear model can cause a linear dose response to appear to have a threshold (*e.g.*, [45]) and consequently can cause bi-linear modeling to predict a

threshold dose, when, in fact, one does not exist. Consequently, the QWG does not recommend log-transforming the experimental doses when applying the bi-linear model.

2.3. Comparison of the different methods to analyze exposure–response relationships & derive points of departure

Each of the above three approaches for analysis of genetic toxicity data in risk assessment was considered by the QWG, and the characteristics, data requirements, advantages and limitations of each are summarized in [Tables 1 and 2](#). In considering the merits and limitations of each approach, the QWG took into account the extensive scientific and practical regulatory experience of the QWG members with these methods, the analyses and recommendations of the Quantitative Working Group of the HESI Genetic Toxicology Testing Committee that have been summarized by Gollapudi et al. [29] and Johnson et al. [30], and additional *de novo* analyses conducted by QWG members. The nature of the analyses and the metrics evaluated are illustrated by Figure 2 from Gollapudi et al. [29], which presents a summary of the quantitative analysis of the dose-response for *hprt* gene mutations induced by MMS *in vitro* in AHH-1 cells.

This figure illustrates the fitted dose–response curves and shows the PoD metrics derived by applying the BMD and bi-linear approaches. In this case, the bilinear model of Lutz and Lutz [64] was employed, although the QWG noted that the R-based segmented analysis described by Muggeo [65] has a number of advantages that make it preferable for bilinear modeling [30].

[Fig. 3](#), which is derived from the results presented by Johnson et al. [30], presents the calculated PoDs from selected examples of data from various *in vitro* and *in vivo* assays for gene mutation and micronucleus induction in animals and mammalian cells exposed to ENU or MNU.

The data summarized in [Figs. 2 and 3](#) exemplify the general findings with a large number of data sets, including those reported by Gollapudi et al. [29] and Johnson et al. [30], those reported in the

Table 2

Advantages and disadvantages of methods considered by the QWG.

Method	BMD	NOGEL	Breakpoint dose determination using a bi-linear model
Measurement	Benchmark dose – dose associated with a specific benchmark response (BMR)	Highest dose with no statistically significant response	Estimate of threshold
Advantages	- Sparse data tends to yield a lower PoD - Uses data efficiently and takes account of the shape of the dose response. (Fits a model to data) - Currently used by many regulatory agencies	- Is easy to apply - Does not require dose response modeling	- Sparse data tends to yield a lower PoD - May be appropriate when mechanistic information supports threshold expectation
Disadvantages	- Requires consensus on appropriate biologically relevant benchmark response (BMR) - Continuous and quantal data are modeled differently	Sparse data tends to yield a higher PoD Statistical assumptions must be met	- Based on specific assumption that data is described by one line of zero slope and another of finite slope - is highly model dependent (other models will fit just as well but predict very different PoD) - is not robust (PoD often cannot be determined for sparse data sets)

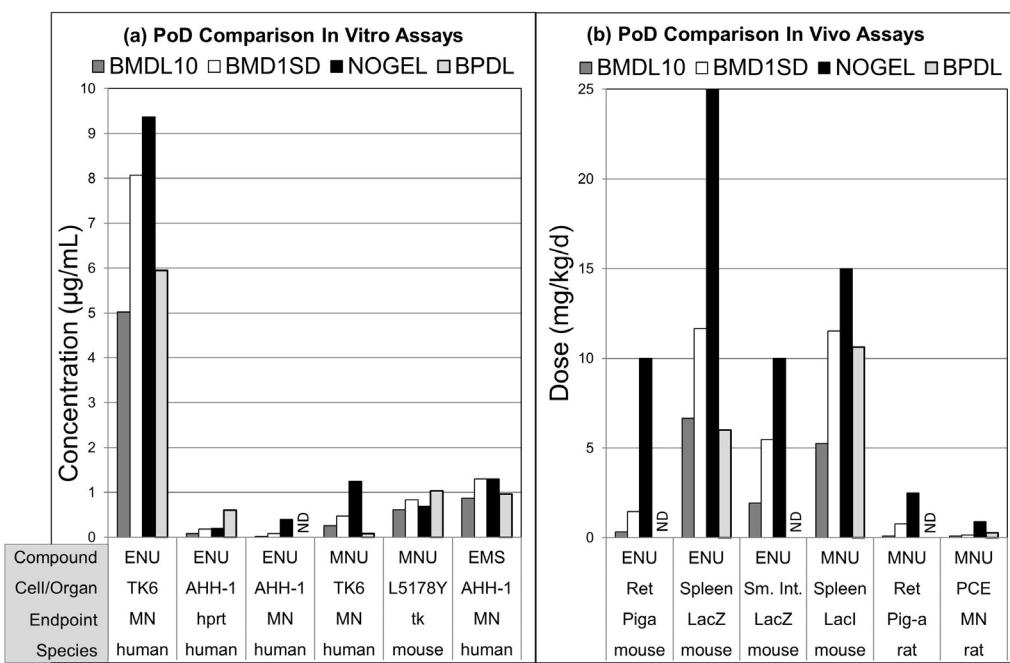


Fig. 3. Comparison of PoD metrics for ENU or MNU response in different assays and endpoints. (a) *In vitro* assays for gene mutation and micronucleus induction. (b) *In vivo* assays for gene mutation in different tissues and micronucleus induction in bone marrow. ND, breakpoint was indeterminate. The PoD metrics presented are the BMDL₁₀, the BMDL_{1SD}, the NOGEL, and the BPDL.

companion report of the IWGT QWG [43], and with the experience of QWG members over the course of their careers.

Noteworthy conclusions and comments, based on critical comparisons of the methods and metrics, include:

- (1) BMD modeling almost always yields a good fit to the data,
- (2) The BMDL₁₀ (the lower confidence limit on the benchmark response rate of 10% over background), derived from the best fitting model, generally provided a conservative (lower) value relative to the NOGEL and the BPDL,
- (3) The BMDL₁₀, though recently employed in studies reported in the literature, is an arbitrary choice of minimal response based on the presumption that a BMR of 10% of the spontaneous rate is a minimal increase in response that is close enough to the range of observable responses to be estimated with reasonable accuracy. Since most genotoxicity assays currently employed do not have the statistical sensitivity to detect less than a doubling of the spontaneous rate (*i.e.*, a 100% increase in the spontaneous rate), the choice of a BMR of 10% results in a PoD that is approximately 1/10 (or less for certain assays) of the detectable NOGEL,
- (4) The BMDL approach is very flexible, can be readily applied to a wide range of datasets, is minimally affected by dose spacing, and requires only three treatment levels. Dose spacing is more critical for determination of the NOGEL and BPD, and BPD analysis requires approximately six treatment groups,
- (5) The BMDL₁₀ and BMDL_{1SD} values are essentially always lower than the NOGEL and are therefore more conservative.

Although a large number of data sets from different types of assays were evaluated, rigorous comparisons were limited to a relatively small sample of structural classes of genotoxins (*i.e.*, mainly monofunctional alkylating agents, aromatic amines, and polycyclic aromatic hydrocarbons). It is therefore desirable to extend these analyses to additional structurally diverse agents. Nonetheless, the QWG considered that a number of conclusions

could be drawn. Below we discuss each approach and provide consensus statements and guidance regarding their use and application.

3. Conclusions and points of consensus

- It is well-accepted that biological understanding of the mechanism of action of some non-DNA-reactive agents supports the expectation of a threshold exposure below which there is no effect.
- Although biological understanding of protective mechanisms and DNA repair capacity often supports the expectation of a no-effect threshold for some DNA-reactive mutagens, it is generally not possible to conclusively establish whether or not a small incremental risk exists within the normal range of spontaneous damage solely by experimental measurement of DNA damage (due to existing background and experimental error) or mathematical analysis of the dose-response data.
- Since the presence of an absolute no-effect threshold cannot be conclusively demonstrated, emphasis should be placed on the determination of PoDs from which limits of exposure associated with an acceptable risk level can be determined. The weight of evidence regarding the mechanism of induction of the damage, extent of repair and other protective cellular defenses, and the available exposure, pharmacokinetic, and metabolism data can then be used to determine if more or less conservative assumptions, uncertainty factors or extrapolation are appropriate in a particular case.
- The rank order of QWG preference for the method of deriving the PoD lies in the order of BMD > NOGEL > the statistical lower bound of the breakpoint dose (BPDL) derived from bilinear modeling. The BMDL is robust and conservative and thus is recommended for general use as the PoD.
- The BMD approach was considered to be the preferred approach for dose(exposure)-response analysis and PoD derivation for

genotoxicity data because: (1) dose–response analysis can be performed on studies with minimal data, (2) it uses the entire data set to derive a BMD estimate, (3) the size of the effect is defined, (4) covariate analysis can be performed, (5) within limits, the PoD value is not adversely affected by experimental design and dose selection (e.g., NOGELs from two different experiments can vary significantly due to differences in dose spacing and statistical sensitivity), (6) confidence limits can be derived. In addition, the BMDL₁₀ was generally the most conservative of the PoD estimates considered by the QWG.

- There was agreement that BMD analysis is suitable for defining PoDs for genetic toxicology data for both discrete (quantal) and continuous responses. Based on experience to date with continuous data, benchmarks defined as a specified increase relative to background appear to have the greatest acceptance with respect to the analysis of genetic toxicology data [60], although other definitions such as the hybrid approach have frequently been used for other types of endpoints [51,57].
- There was agreement that the NOGEL can be a suitable PoD for genetic toxicity dose–response data. Appropriate statistical criteria for deriving and applying NOGEL values are necessary, as is the case for all approaches. The ILSI-HESI GTTC QAW standard operating procedure [68] and R package [67] are suitable for NOGEL analyses using Dunnett's, Dunn's or Dunnett's T3 tests depending on the distributional characteristics of the data [30]. A trend test can also be used for defining a NOGEL, especially for studies with large numbers of dose groups; however software and methods for performing these tests are not yet included in the R package and SOP. Basing a NOGEL on a stepwise trend test that uses all the data may increase the statistical power of the NOGEL determination.
- Although there was general agreement among the QWG that the bi-linear model could be used to determine a PoD when mechanistic data supports the existence of a threshold dose, significant drawbacks of the BPD method were recognized. First, the assumption that the true model is composed of two intersecting straight lines is not appropriate unless there is mechanistic information that supports it. The PoD (i.e., BPD, or threshold dose estimate) obtained from this approach is highly dependent upon the validity of this assumption. For example, Crump [45] showed that there will always be a dose–response model that fits the data as well as the bi-linear and that predicts a zero threshold dose (i.e., that has a positive slope at all doses in the low-dose range). This underlies the reason for selecting a BMR in the benchmark approach that is greater than zero – this ensures that the BMD is always greater than zero (and therefore useful as a PoD). Secondly, bi-linear modeling cannot be used to determine a PoD in all cases, even in all cases where the data are from an experiment where a threshold is deemed to be plausible. This is a result of the fairly frequent finding that the statistical lower bound on the threshold estimate will include zero and consequently be unsuitable as a PoD. Furthermore, the QWG considered the bilinear model to be more dependent on factors such as the number and spacing of doses and statistical power within dose groups than other approaches. Therefore, the QWG considered the use of the bilinear model to be too restricted. In contrast, the BMD approach is generally applicable irrespective of whether or not there is a threshold dose, and also has other advantages described above.
- In general, the best fitting dose–response model consistent with the biological mechanistic information for a given data set should be used to analyze the data within the observable dose range and to derive a PoD.
- In all cases, there is a need for high quality data, with transparent characterization of the uncertainties around the risk estimation approach.

Tables 1 and 2 summarize some of the recommended data needs and characteristics of the three approaches for computing PoDs, and the major advantages and disadvantages of each.

In summary, the consensus among the QWG participants was that determination of the PoD by the BMD method is favored whenever it is supported by the data; otherwise the NOGEL method may be used. The QWG was divided regarding the use of bi-linear modeling for determining a PoD. Most QWG members agreed that the bi-linear model could be used for determining a PoD when mechanistic data support the existence of a threshold dose, while at least one panel member disagreed, citing the disadvantages of the bi-linear method discussed above. Thus, each of the 3 approaches can be used to define a PoD, but the preference of the QWG lies in the order: BMD > NOGEL > bilinear.

Considerations when extrapolating dose–response data from PoDs to determine low-dose human risks are summarized in the second publication in this series [43]. These factors should be determined on a case-by-case basis by regulatory authorities and risk managers within their areas of responsibility, but the QWG provides a general framework of factors to consider and approaches for incorporating them into regulatory decision-making. Uncertainties include those in extrapolating below the PoD, in extrapolating across test systems and from laboratory results to human effects, and human inter-individual variability and genetic susceptibility. These uncertainties determine the magnitude of safety and uncertainty factors that must be applied to a PoD to determine an estimate of acceptable exposure.

Contributions of authors

All authors participated in the discussions that led to the consensus conclusions and recommendations reported herein, and all authors contributed to the writing and editing of this report. J.T.M. served as Chair of the QWG and played a major role in organizing and leading the group discussions, as well as drafting significant sections of this publication. The roles of Co-Chair and Rapporteur of the QWG were initially held by V.T. and M.J.S., respectively, but due to their inability to attend the meetings at Foz do Iguaçu, R.F. and P.A.W., respectively, assumed these roles. Major sections of the manuscript were initially drafted by G.E.J. and K.S.C. (methods of dose–response analysis and derivation of PoDs) and by J.T.M., V.T., and R.F. (introduction and objectives). Many QWG members participated in small sub-groups that formulated issues to be addressed and developed proposed consensus points and recommendations within sub-topic areas. Sub-topic group leaders were J.T.M., V.T., R.F. (need for quantitative approaches, background, objectives), G.E.J., K.S.C. (methods to analyze dose–response data, descriptive metrics, and PoD derivation), D.A.E., R.S. (threshold issues and applications of PoDs), and J.T.M., R.F., P.A.W. (overall recommendations). G.E.J., L.H.-S., and K.S.C. performed the dose–response analyses that were conducted by the QWG. D.D.L. and T.M. managed the reference formatting and citations.

Conflict of interest statement

J.T.M. consults for regulated industries, government agencies, and laboratories that develop and/or perform regulatory genetic toxicology tests. No other conflicts of interest are noted.

Acknowledgements

We acknowledge support from the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs, contract number NC/K500033/1) for Dr. Johnson. Funding for the development of the G4 genetic toxicity dose–response

database described by Gollapudi et al. [29] and Johnson et al. [30] and of R-based procedures for derivation of genetic toxicology PoDs (i.e., drsmooth), and for support of open access of this publication, was provided by Health Canada under the Chemicals Management Plan Research Fund. We thank the ILSI/HESI Quantitative Working Group for sharing data and analyses collected and employed by that Working Group, and note that their analyses were a critical component of our deliberations.

Appendix 1. List of abbreviations and definitions¹

BMC (Benchmark Concentration): The concentration of a substance that is associated with a specified low incidence of response of a biological effect, generally in the range of 1–10%.

BMCL (Benchmark Concentration Low): A lower one-sided confidence limit on the BMC.

BMD (Benchmark Dose; for quantal data): A dose or exposure of a substance associated with a specified absolute increase in the incidence of a biological response, generally in the range of 1–10%.

BMD (Benchmark Dose; for continuous data): A dose or exposure that produces a predetermined percentage change in response rate of an adverse effect relative to the existing background incidence, generally in the range of 1–10% increase in the background.

BMD₁₀ (Benchmark Dose₁₀): Benchmark Dose associated with a 10% response adjusted for (for quantal data) or relative to (for continuous data) background.

BMD1SD: Benchmark dose associated with an increase relative to background of 1SD of the mean background value.

BMDL (Benchmark Dose Lower Limit): A statistical lower confidence limit on the dose or concentration at the BMD.

BMDL₁₀ (Benchmark Dose Lower Limit₁₀): The lower confidence limit of a benchmark response rate of 10% (for quantal data) or 10% increase in the background frequency (for continuous data).

BMR (Benchmark Response): The response, generally expressed as in excess of background, at which a benchmark dose or concentration is desired (see Benchmark Dose, Benchmark Concentration).

BPD (Breakpoint Dose): The dose at which the slope changes from zero (horizontal) to positive, with its standard error forming the confidence bounds [30].

BPDL: Lower confidence limit on the breakpoint dose.

CHMP: Committee for Medicinal Products for Human Use (CHMP), a committee of the European Medicines Agency.

DHEW: U. S. Department of Health, Education, and Welfare.

EFSA: European Food Safety Authority.

EMA: European Medicines Agency.

EMS: Ethyl Methanesulfonate.

ENU: N-Nitroso-N-methylurea, Ethyl Nitrosourea.

FDA: U.S. Food and Drug Administration.

HHRA: Human Health Risk Assessment.

ICH: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use.

ICH S2(R1): ICH guideline S2 (R1) on genotoxicity testing and data interpretation for pharmaceuticals intended for human use.

ILSI-HESI GTTC QAW: International Life Sciences Institute, Health and Environmental Sciences Institute, Genetic Toxicology Testing Committee, Quantitative Assessment Working Group.

IWGT: International Workshops on Genotoxicity Testing.

¹ See also: <http://www.epa.gov/ncea/bmds/bmds.training/appendices/glossary.htm>, Benchmark Dose Software (BMDS), Appendix C: Glossary of Terms, APPENDIX C, GLOSSARY OF TERMS: <http://ofmpub.epa.gov/sor.internet/registry/termreg/searchandretrieve/glossariesandkeywordlists/search.do>, Vocabulary Catalog List Detail – Integrated Risk Information System (IRIS) Glossary, Terms & Acronyms; Gollapudi et al., EMM, 54, 8–18, 2013.

LOGEL (Lowest Observed Genotoxic Effect Level) The lowest tested dose for which a statistically significant increase in the incidence of the genotoxic effect is observed, relative to an appropriate background control incidence.

MMS: Methyl Methanesulfonate.

MN: micronucleus.

MNU: N-Nitroso-N-methylurea, Methyl Nitrosourea.

NOAEL (No Observed Adverse Effect Level): The highest tested dose for which no statistically significant increase in the incidence of an adverse effect is observed and below which no statistically significant response occurs relative to an appropriate untreated control.

NOGEL (No Observed Genotoxic Effect Level): The highest tested dose for which no statistically significant increase in the incidence of the genotoxic effect is observed, and below which no statistically significant response occurs, relative to an appropriate untreated control (i.e., background).

PCE: polychromatic erythrocyte.

PoD (Point of Departure): The point on a dose–response curve established from experimental data from which extrapolation below the PoD may be employed, in conjunction with the application of uncertainty factors, for low-dose risk assessment and determination of an acceptable exposure level, or reference dose. A PoD can be a data point or an estimated point that is derived from observed dose–response data. This point can be the lower bound on dose for an estimated incidence or a change in response level from a dose–response model (BMD), or a NOAEL or LOAEL for an observed incidence, or change in level of response.

QWG: IWGT Working Group on Quantitative Approaches to Genetic Toxicology Risk Assessment.

Td: Threshold Effect Level is the dose at which the slope changes from zero (horizontal) to positive when using the bilinear hockey stick model [29].

Td CI: The confidence limits on the Td.

TdL CI: The lower confidence limit on the Td.

U.S. EPA: United States Environmental Protection Agency.

References

- [1] L. Müller, D. Blakey, K.L. Dearfield, S. Galloway, P. Guzzie, M. Hayashi, P. Kasper, D. Kirkland, J.T. MacGregor, J.M. Parry, L. Schechtman, A. Smith, N. Tanaka, D. Tweats, H. Yamasaki, Strategy for genotoxicity testing and stratification of genotoxicity test results – report on initial activities of the IWGT Expert Group, *Mutat. Res.* 540 (2003) 177–181.
- [2] D.J. Kirkland, M. Hayashi, D. Jacobson-Kram, P. Kasper, J.T. MacGregor, L. Müller, Y. Uno, Summary of major conclusions from the 4th IWGT, San Francisco, 9–10 September, 2005, *Mutat. Res.* 627 (2007) 5–9.
- [3] V. Thybaud, J.T. MacGregor, L. Müller, R. Cebelli, K. Dearfield, G. Douglas, P.B. Farmer, E. Gocke, M. Hayashi, D.P. Lovell, W.K. Lutz, D. Marzin, M. Moore, T. Nohmi, D.H. Phillips, B.J. van Benthem, Strategies in case of positive in vivo results in genotoxicity testing, *Mutat. Res.* 723 (2011) 121–128.
- [4] J.S. Wassom, H.V. Malling, K. Sankaranarayanan, P.Y. Lu, Reflections on the origins and evolution of genetic toxicology and the Environmental Mutagen Society, *Environ. Mol. Mutagen.* 51 (2010) 746–760.
- [5] Ministry of Labour, Japan, Standards for Mutagenicity Tests using Microorganisms, Notification based on Article 57-2-1 of the Industrial Safety and Health Law, Ordinance No. 57, 1972.
- [6] Department of National Health and Welfare and the Department of Environment Advisory Committee on Mutagenesis, Guidelines on the use of mutagenicity tests in the toxicological evaluation of chemicals, *Environ. Mol. Mutagen.* 11 (1988) 261–304.
- [7] European Union Council, Council Directive 83/570/EEC of 26 October 1983 amending Directives 65/65/EEC, 75/318/EEC and 75/319/EEC on the approximation of provisions laid down by law, regulation or administrative action relating to proprietary medicinal products, 1983.
- [8] U.S. Environmental Protection Agency, Toxic substances control act (1976), in: A. Hollaender, F.J. de Serres (Eds.), *Chemical Mutagens: Principles and Methods for Their Detection*, Plenum Press, New York, 1978, pp. 287–335.
- [9] U.S. Environmental Protection Agency, Subdivision F: Toxicology Data Requirements, U.S. Code of Federal Regulations 40, 2005.
- [10] W.G. Flamm, L.R. Valcovic, F.J. de Serres, W. D'Aguanno, L. Fishbein, S. Green, H.V. Malling, V. Mayer, M.J. Prival, G. Wolff, E. Zeiger, Approaches to determining the mutagenic properties of chemicals: risk to future generations, *J. Environ. Toxicol. Chem.* 1 (1977) 301–352.

[11] B.N. Ames, Carcinogens are mutagens: their detection and classification, *Environ. Health Perspect.* 6 (1973) 115–118.

[12] J. McCann, E. Choi, E. Yamasaki, B.N. Ames, Detection of carcinogens as mutagens in the *Salmonella*/microsome test: assay of 300 chemicals, *Proc. Natl. Acad. Sci. U. S. A.* 72 (1975) 5135–5139.

[13] B.A. Fetterman, B.S. Kim, B.H. Margolin, J.S. Schildcrout, M.G. Smith, S.M. Wagner, E. Zeiger, Predicting rodent carcinogenicity from mutagenic potency measured in the Ames *Salmonella* assay, *Environ. Mol. Mutagen.* 29 (1997) 312–322.

[14] J.S. Schildcrout, B.H. Margolin, E. Zeiger, Predicting rodent carcinogenicity using potency measures of the *in vitro* sister chromatid exchange and chromosome aberration assays, *Environ. Mol. Mutagen.* 33 (1999) 59–64.

[15] K.L. Witt, A. Knapton, C.M. Wehr, G.J. Hook, J. Mirsalis, M.D. Shelby, J.T. MacGregor, Micronucleated erythrocyte frequency in peripheral blood of B6C3F(1) mice from short-term, prechronic, and chronic studies of the NTP carcinogenesis bioassay program, *Environ. Mol. Mutagen.* 36 (2000) 163–194.

[16] D.A. Eastmond, A. Hartwig, D. Anderson, W.A. Anwar, M.C. Cimino, I. Dobrev, G.R. Douglas, T. Nohmi, D.H. Phillips, C. Vickers, Mutagenicity testing for chemical risk assessment: update of the WHO/IPCS harmonized scheme, *Mutagenesis* 24 (2009) 341–349.

[17] R.D. Snyder, J.W. Green, A review of the genotoxicity of marketed pharmaceuticals, *Mutat. Res.* 488 (2001) 151–169.

[18] T. Sofuni, M. Hayashi, T. Nohmi, A. Matsuoka, M. Yamada, E. Kamata, Semiquantitative evaluation of genotoxic activity of chemical substances and evidence for a biological threshold of genotoxic activity, *Mutat. Res.* 464 (2000) 97–104.

[19] D. Kirkland, M. Aardema, L. Henderson, L. Müller, Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity, *Mutat. Res.* 584 (2005) 1–256.

[20] R.K. Elespuru, R. Agarwal, A.H. Atrakchi, C.A. Bigger, R.H. Heflich, D.R. Jagan-nath, D.D. Levy, M.M. Moore, Y. Ouyang, T.W. Robison, R.E. Sotomayor, M.C. Cimino, K.L. Dearfield, Current and future application of genetic toxicity assays: the role and value of *in vitro* mammalian assays, *Toxicol. Sci.* 109 (2009) 172–179.

[21] J.J. Clarke, T.E. Lawlor, W. Madraymootoo, K. Pant, R.R. Young, V.O. Wagner III, M.J. Aardema, Summary of *in vitro* genetic toxicology assay results: expected and unexpected effects of recent study design modifications, *Environ. Mol. Mutagen.* 53 (2012) 631–635.

[22] M.D. Fellows, S. Boyer, M.R. O'Donovan, The incidence of positive results in the mouse lymphoma TK assay (MLA) in pharmaceutical screening and their prediction by MultiCase MC4PC, *Mutagenesis* 26 (2011) 529–532.

[23] U.S. Food and Drug Administration, Recommended Approaches to Integration of Genetic Toxicology Study Results, Center for Drug Evaluation and Research Guidance for Industry and Review Staff, 2006, <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM079257.pdf> (accessed 08.28.14).

[24] International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH harmonized tripartite guideline guidance on genotoxicity testing and data interpretation for pharmaceuticals intended for human use S2(R1), 77 FR 110, 2012, 33748–33749.

[25] European Medicines Agency, Guideline on limits of genotoxic impurities, EMEA/CHMP/QWP/251344/2006, 2006, http://www.ema.europa.eu/ema/pages/includes/document/open_document.jsp?webContentId=WC500002903 (accessed 08.28.14).

[26] U.S. Food and Drug Administration Guidance for Industry: Genotoxic and Carcinogenic Impurities in Drug Substances and Products: Recommended Approaches, Center for Drug Evaluation and Research, 2008, <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM079235.pdf> (accessed 08.28.14).

[27] International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Guideline M7 on assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk – Step 3, 2013, http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/02/WC500139217.pdf (accessed 08.28.14).

[28] V. Thybaud, M. Aardema, D. Casciano, V. Dellarco, M.R. Embry, B.B. Gollapudi, M. Hayashi, M.P. Holsapple, D. Jacobson-Kram, P. Kasper, J.T. MacGregor, R. Rees, Relevance and follow-up of positive results in *in vitro* genetic toxicity assays: an ILSI-HESI initiative, *Mutat. Res.* 633 (2007) 67–79.

[29] B.B. Gollapudi, G.E. Johnson, L.G. Hernandez, L.H. Pottenger, K.L. Dearfield, A.M. Jeffrey, E. Julien, J.H. Kim, D.P. Lovell, J.T. MacGregor, M.M. Moore, B.J. van Benthem, P.A. White, E. Zeiger, V. Thybaud, Quantitative approaches for assessing dose–response relationships in genetic toxicology studies, *Environ. Mol. Mutagen.* 54 (2013) 8–18.

[30] G.E. Johnson, L. Soeteman-Hernandez, B.B. Gollapudi, O.G. Bodger, K.L. Dearfield, R.L. Heflich, J.G. Hixon, D.P. Lovell, J.T. MacGregor, L.H. Pottenger, C.M. Thompson, L. Abraham, V. Thybaud, J.Y. Tanir, E. Zeiger, B.J. van Benthem, P.A. White, Derivation of points of departure (PoD) estimates in genetic toxicology studies and their potential application in risk assessment, *Environ. Mol. Mutagen.* 55 (2014) 609–623.

[31] M. Guérard, M. Baum, A. Bitsch, G. Eisenbrand, A. Elhajouji, B. Epe, M. Haberbeyler, B. Kaina, H.J. Martus, S. Pfuhler, Z. Schmitz, A. Sutter, A.D.Z.C. Thomas, R. Froetschl, Assessment of mechanisms driving non-linear dose-response relationships in genotoxicity testing (a report of the Gesellschaft für Umwelt-Mutationsforschung e.V. (GUM; Society for Environmental Mutation Research) expert committee on thresholds for genotoxicity), *Mutat. Res.* (2014) (under revision).

[32] D.J. Brusick, *Methods for Genetic Risk Assessment*, CRC Press, Boca Raton, FL, 1994.

[33] J. Favor, D. Layton, G. Segal, J. Wassom, J. Burkhardt, G. Douglas, K. Dearfield, D. Brusick, International Commission for Protection Against Environmental Mutagens and Carcinogens, Genetic risk extrapolation from animal data to human disease. A Taskgroup Report. No. 00105, *Mutat. Res.* 330 (1995) 23–34.

[34] National Academy of Sciences (US), *Risk Assessment in the Federal Government: Managing the Process*, National Academy Press, Washington, DC, 1983.

[35] J. Lewtas, D.M. DeMarini, J. Favor, D.W. Layton, J.T. MacGregor, J. Ashby, P.H.M. Lohman, R.H. Haynes, M.L. Mendelsohn, Risk characterization strategies for genotoxic environmental agents, in: D.J. Brusick (Ed.), *Methods for Genotoxic Risk Assessment*, CRC Press, Boca Raton, FL, 1994, pp. 125–169.

[36] U.S. Food and Drug Administration *Toxicological Principles for the Safety Assessment of Food ingredients*, Redbook 2000 (Revised 2007). Center for Food Safety and Applied Nutrition, 2007, <http://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/IngredientsAdditivesGRASPackaging/ucm2006826.htm> (accessed 08.28.14).

[37] M.A. Cheeseman, E.J. Machuga, A.B. Bailey, A tiered approach to threshold of regulation, *Food Chem. Toxicol.* 37 (1999) 387–412.

[38] A.R. Rulis, De Minimus and the threshold of regulation, in: C.W. Felix (Ed.), *Food Protection Technology*, Lewis Publishers, Inc., Chelsea, MI, 1986, pp. 29–37.

[39] Health Canada, The assessment of mutagenicity. Health Protection Branch mutagenicity guidelines, *Environ. Mol. Mutagen.* 21 (1993) 15–37.

[40] A. Elhajouji, M. Lukamowicz, Z. Cammerer, M. Kirsch-Volders, Potential thresholds for genotoxic effects by micronucleus scoring, *Mutagenesis* 26 (2011) 199–204.

[41] A. Lynch, J. Harvey, M. Aylott, E. Nicholas, M. Burman, A. Siddiqui, S. Walker, R. Rees, Investigations into the concept of a threshold for topoisomerase inhibitor-induced clastogenicity, *Mutagenesis* 18 (2003) 345–353.

[42] H.M. Bolt, H. Foth, J.G. Hengstler, G.H. Degen, Carcinogenicity categorization of chemicals-new aspects to be considered in a European perspective, *Toxicol. Lett.* 151 (2004) 29–41.

[43] J.T. MacGregor, R. Frötschl, P.A. White, K. Crump, D.A. Eastmond, S. Fukushima, M. Guérard, M. Hayashi, L. Soeteman-Hernandez, G.E. Johnson, T. Kasamatsu, D.D. Levy, T. Morita, L. Müller, R. Schoeny, M. Schuler, V. Thybaud, IWGT Report on Quantitative Approaches to Genotoxicity Risk Assessment II. Use of Point-of-Departure (PoD) metrics in defining acceptable exposure limits and assessing human risk, *Mutat. Res.* 783 (2015) 66–78.

[44] U.S. Environmental Protection Agency, *Guidelines for Cancer Risk Assessment*, 2005, <http://www.epa.gov/cancerguidelines/> (accessed 08.28.14).

[45] K.S. Crump, Use of threshold and mode of action in risk assessment, *Crit. Rev. Toxicol.* 41 (2011) 637–650.

[46] Health Canada, *Health-based Tolerable Daily Intakes/Concentrations and Tumorigenic Doses/Concentrations for Priority Substances*, Environmental Health Directorate, Health Protection Branch, Ottawa, 1996.

[47] U.S. Food and Drug Administration *Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers*, Center for Drug Evaluation and Research, 2005, <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm065014.htm> (accessed 04.28.14).

[48] K.S. Crump, A new method for determining allowable daily intakes, *Fundam. Appl. Toxicol.* 4 (1984) 854–871.

[49] L.G. Hernandez, W. Slob, S.H. van, B.J. van Benthem, Can carcinogenic potency be predicted from *in vivo* genotoxicity data? A meta-analysis of historical data, *Environ. Mol. Mutagen.* 52 (2011) 518–528.

[50] L.G. Hernandez, B.J. van Benthem, W. Slob, Estimating the carcinogenic potency of chemicals from the *in vivo* micronucleus test, in: RIVM Report 340700007/2012, Bilthoven, 2012.

[51] National Academy of Sciences (US), *Toxicological Effects of Methylmercury*, National Academy Press, Washington, DC, 2000.

[52] U.S. Environmental Protection Agency, *Health Assessment of 1,3-Butadiene*, EPA EPA/600/P-98/001F, 2002, <http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=5449#Download> (accessed 08.28.14).

[53] EFSA Scientific Committee, Guidance of the Scientific Committee on a request from EFSA on the use of the benchmark dose approach in risk assessment, EFSA J. 1150 (2009) 1–72, <http://www.efsa.europa.eu/en/efsajournal/pub/1150.htm> (accessed 08.28.14).

[54] D.W. Gaylor, W. Slikker Jr., Risk assessment for neurotoxic effects, *Neurotoxicology* 11 (1990) 211–218.

[55] K. Crump, J. Viren, A. Silvers, H. Clewell III, J. Gearhart, A. Shipp, Reanalysis of dose–response data from the Iraqi methylmercury poisoning episode, *Risk Anal.* 15 (1995) 523–532.

[56] E. Budtz-Jorgensen, N. Keiding, P. Grandjean, Benchmark dose calculation from epidemiological data, *Biometrics* 57 (2001) 698–706.

[57] E. Budtz-Jorgensen, D. Bellinger, B. Lanphear, P. Grandjean, An international pooled analysis for obtaining a benchmark dose for environmental lead exposure in children, *Risk Anal.* 33 (2013) 450–461.

[58] K. Crump, Critical issues in benchmark calculations from continuous data, *Crit. Rev. Toxicol.* 32 (2002) 133–153.

[59] R.W. West, R.L. Kodel, Statistical methods of risk assessment for continuous variables, *Comm. Stat. Theor. Methods* 22 (1993) 3363–3376.

- [60] W. Slob, R.W. Setzer, Shape and steepness of toxicological dose-response relationships of continuous endpoints, *Crit. Rev. Toxicol.* 44 (2014) 270–297.
- [61] W. Slob, Dose-response modeling of continuous endpoints, *Toxicol. Sci.* 66 (2002) 298–312.
- [62] J.A. Davis, J.S. Gift, Q.J. Zhao, Introduction to benchmark dose methods and U.S. EPA's benchmark dose software (BMDS) version 2.1.1, *Toxicol. Appl. Pharmacol.* 254 (2011) 181–191.
- [63] European Medicines Agency, Studies Assessed by the European Medicines Agency Indicate No Increased Risk of Developing Cancer for Patients Who Have Taken Viracept Contaminated With Ethyl Mesilate, Press Release, Doc. Ref. EMEA/CHMP/382256/2008, London, July, 2008, www.ema.europa.eu/docs/en_GB/document_library/Press_release/2009/11/WC500015040.pdf (accessed 08.28.14).
- [64] W.K. Lutz, R.W. Lutz, Statistical model to estimate a threshold dose and its confidence limits for the analysis of sublinear dose-response relationships, exemplified for mutagenicity data, *Mutat. Res.* 678 (2009) 118–122.
- [65] V. Muggeo, Segmented: An R Package to Fit Regression Models With Broken-line Relationships, *R News*, 2008, pp. 20–25, <http://cran.r-project.org/doc/Rnews/> (accessed 08.28.14).
- [66] E. Gocke, M. Wall, In vivo genotoxicity of EMS: statistical assessment of the dose response curves, *Toxicol. Lett.* 190 (2009) 298–302.
- [67] G. Hixon, A. Bichteler, Drssmooth: dose-response modeling with smoothing splines, 2014, <http://CRAN.R-project.org/package=drsmooth> (accessed 08.28.14).
- [68] C.M. Thompson, L. Abraham, A. Bichteler, v Hixon, Deriving point of departure estimates for genotoxicity datasets using BMD modeling and R-based packages, Health Canada, Ottawa, Ontario, 2014.